The Wyner-Ziv approach to distributed source coding

Deepak Sridhara

February 6, 2003

1 Introduction

In a communication network, suppose there are a collection of (correlated) *source* nodes that want to communicate to a common *sink* node, then the distributed source coding problem is one of designing a system (of encoders and decoder) that optimally compresses the correlated sources. The key aspect of the design is that the source nodes aren't allowed to communicate with one another –i.e., the sources are encoded separately, but are decoded jointly. (If they could communicate, then the problem reduces to the single user source coding problem.) The goal of the distributed source coding problem is to design a system that approaches the optimal performance that can be obtained if the source nodes could communicate with one another.

Slepian and Wolf (SW) showed that in the case of two discrete sources X and Y, it is *theoretically* possible to achieve the same level of performance in a system where the encoders for X and Y are separated as in a system where they are not. More formally, suppose X and Y are two statistically dependent discrete random processes, taking values in finite alphabets[1], which are encoded by two separate encoders, but are decoded by a joint decoder. Then the SW result says that even if the encoders are independent, to achieve reliable transmission of both X and Y, the rates of transmission must be such that $R_X \geq H(X|Y)$, $R_Y \geq H(Y|X)$, and $R_X + R_Y \geq H(X,Y)$.

Wyner considered the following specialization to the above problem: suppose the decoder is primarily interested in reconstructing X, and Y is considered as some side-information on X available to the decoder. Then the problem specializes to the optimal rates of sending X and Y so that the decoder can reconstruct X with high reliability. The design of such a system will be called the zero-error source coding problem with side information [1], since the system is to be designed such that there is an arbitrarily small probability of error in the reconstruction of X. (If Y is transmitted at a rate $R_Y \geq H(Y)$, then from the SW result, it must be theoretically possible to transmit X at a rate $R_X = H(X|Y) + \epsilon$, for ϵ arbitrarily small.) The problem is stated more formally in Section 2.

A counterpart to the above problem is the lossy source coding problem, introduced by Wyner and Ziv [2]. Suppose Y is readily available to the decoder, then a system must be designed that encodes X at the optimum rate so that X can be reconstructed within a certain distortion (or, fidelity criterion) by the decoder. Wyner and Ziv computed the optimum rate $R_Y^*(d)$ of transmitting X for the decoder to reconstruct $\hat{X}$ that is within an average distortion d of X (for a specified distortion measure). The

[1] T. M. Cover extended this result to the case of ergodic sources, [*IEEE Trans. on Info. Thry.*, Mar 1975].

parameter $R_Y^*(D)$ is called the rate distortion function with side information Y. The design of such a system will be called the lossy source coding problem with side information.

Theoretical limits of communication have been computed for the above cases. However until now, there has been relatively very little work on explicitly designing systems that approach these limits. The proofs showing the achievability of the limits are based entirely on random coding arguments. A few recent works have considered the case of lossy source coding with side information at the decoder [3], [4]. A toy example in [3] is chosen to illustrate how side information at the decoder can help in reducing the transmission rate (or, in improving the compression).

The two problems along with the results are stated formally in Section 2 and Section 3. An example to illustrate the principle behind source coding with side information is presented in Section 4. Finally, an unsolved problem in this area is stated.

2 Zero-error source coding with side information

Let X and V be two discrete random variables that take values in two finite alphabets $\mathcal{X}$ and $\mathcal{V}$ respectively. Let $\{X_i, V_i\}_{i=1}^n$ be a sequence of n independent copies of (X, V). Consider the scenario described in Figure 1.

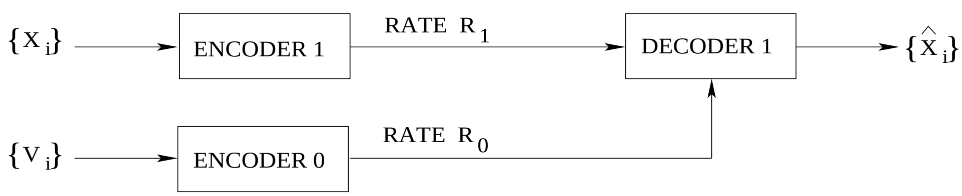

Figure 1: Source coding with side information: one source.

The goal is to deliver a reliable reproduction $\{\hat{X}_k\}$ of $\{X_k\}$ to receiver 1. Encoders 0 and 1 are separated, whereas the decoder can observe both the encoded sequences. The sequence $\{V_k\}$ that is statistically dependent on $\{X_k\}$ is used as side information at the decoder. Encoder i sends information at a rate of R_i bits/symbol. A rate pair (R_0, R_1) is said to be achievable if it is possible (in an information theoretic sense) to build a system (of encoders and decoder) as in Figure 1, having parameters R_0 and R_1, with arbitrarily high reliability – i.e., if the sequences are encoded in blocks of n, there exists mappings

$$\begin{array}{ll} \text{Encoder 0:} & f_0 : \mathcal{V}^n \to \{0, 1, \ldots, 2^{nR_0} - 1\}, \\ \text{Encoder 1:} & f_1 : \mathcal{X}^n \to \{0, 1, \ldots, 2^{nR_1} - 1\}, \\ \text{Decoder:} & g_1 : \{0, 1, \ldots, 2^{nR_0} - 1\} \times \{0, 1, \ldots, 2^{nR_1} - 1\} \to \mathcal{X}^n \end{array}$$

such that the reconstruction error[2] $\Delta = \frac{1}{n} E[d_H(\mathbf{X}^n, \hat{\mathbf{X}}^n)]$, can be made arbitrarily small. (Note that the above mappings define a code $(n, 2^{nR_0}, 2^{nR_1}, \Delta)$.) The source coding with side information problem is to determine the set $\mathscr{R}$ of all achievable rate pairs.

The result is stated as follows: if the joint distribution of X and V is given by $Q(x, v) = Pr\{X = x, V = v\}$, then all rate pairs (R_0, R_1) are achievable if and only if:

$$R_1 \geq H(X|W) \text{ and } R_0 \geq I(V; W)$$

[2]In this case, the reconstruction error is measured in terms of the Hamming metric $d_H(\cdot, \cdot)$.

for some auxiliary random variable W satisfying

$$\text{(a)} \sum_{w \in \mathcal{W}} p(x, v, w) = Q(x, v), \quad \text{(b)}\ p(x, v, w) = Q(x, v) p_t(w|v)$$

In [1], Wyner considered a more general structure with three encoders and two decoders as shown in Figure 2, where the goal was to reconstruct sources X and Y reliably given the side-information from V. (The sequence $\{V_k\}$ is statistically dependent on $\{X_k\}$ and $\{Y_k\}$.) Wyner determined the set $\mathscr{R} = \{(R_0, R_1, R_2) \mid (R_0, R_1, R_2) \text{ is achievable}\}$ for this problem. The structure of Figure 1 is a special case of the above design where $Y \equiv 0$; another case of interest is when $V_k = (X_k, Y_k)$. We will discuss the result and the proof for the special case $Y \equiv 0$.

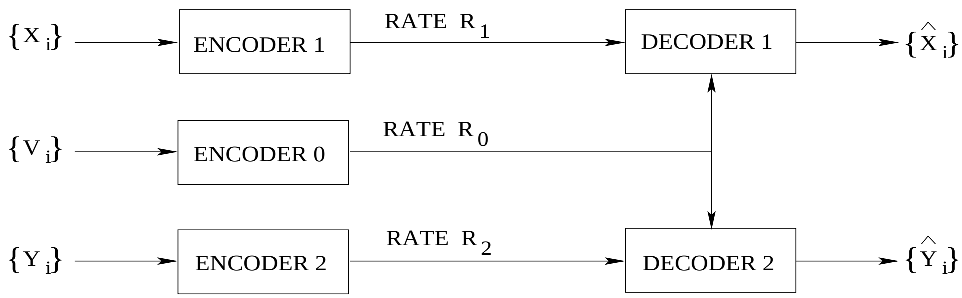

Figure 2: Source coding with side information: two sources (Wyner design).

3 Lossy source coding with side information

Let $\{X_k, Y_k\}_{k=1}^{\infty}$ be a sequence of independent copies of two statistically dependent random variables (X, Y). As in the previous case, X and Y take values in two finite sets $\mathcal{X}$ and $\mathcal{Y}$. The sequence $\{X_k\}$ is encoded in blocks of length n to a rate of R bits/source-symbol, from which the decoder reconstructs the sequence $\{\hat{X}_k\}$. (Let the reconstructed symbols $\hat{x_k}$ take values in $\hat{\mathcal{X}}$.) The average distortion in the reconstructed sequence is $(1/n) \sum_{k=1}^{n} E[D(X_k, \hat{X}_k)]$, where $D(x, \hat{x}) \geq 0$ for $x \in \mathcal{X}$ and $\hat{x} \in \hat{\mathcal{X}}$ is a specified distortion measure. Consider the scenario in Figure 3. An encoder and decoder for the system in Figure 3 must be designed so that the encoder can transmit at a rate R and the decoder can reconstruct the signal within an average distortion d, using the available side-information. The lossy source coding problem with side information asks for the minimum allowable rate of transmission in this situation. Two cases are possible (as shown in the figure) – one, where the encoder has access to the side information and the other, where it does not. Let $R_{X|Y}(d)$ be the smallest allowable rate in the former case, and let $R_Y^*(d)$ be the smallest allowable rate in the latter. Wyner and Ziv give expressions

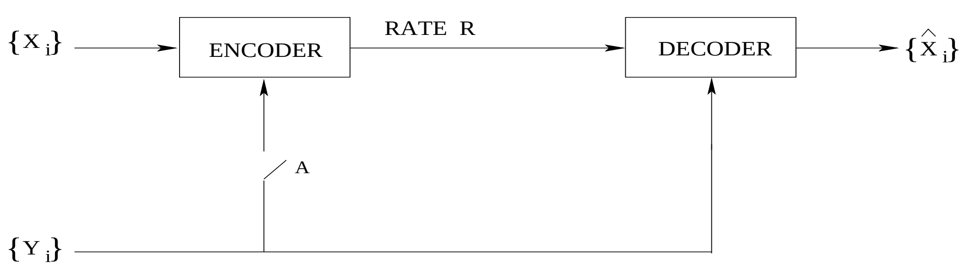

Figure 3: Rate distortion with side information at the decoder.

to compute these quantities and show that in general $R^*_Y(d) \geq R_{X|Y}(d)$ –i.e., the transmission rate can be reduced if the encoder also has access to the side information. Wyner extended this result to more general sources (that are not necessarily discrete) [5]. Of particular interest is the case when X and Y are jointly Gaussian. In this case, it can be shown that $R^*_Y(d) = R_{X|Y}(d)$ – a result similar to the Slepian-Wolf result for lossless coding of X given side-information Y, i.e., it says that the transmission rate cannot be lowered even if the encoder has access to the side information.

A pair (R, d) is said to be achievable if it is possible (in an information theoretic sense) to build a system as in Figure 3, having a transmission rate of R and an average distortion arbitrarily close to d – i.e., if the sequences are encoded in blocks of n, there exists mappings

$$\text{Encoder:} \qquad f_E : \mathcal{X}^n \to \{0, 1, \ldots, 2^{n(R)} - 1\},$$

$$\text{Decoder:} \quad f_D : \mathcal{Y}^n \times \{0, 1, \ldots, 2^{n(R)} - 1\} \to \hat{\mathcal{X}}^n,$$

such that:

$$\limsup_{n\to\infty} E[D(\mathbf{X}^n, f_D(\mathbf{Y}^n, f_E(\mathbf{X}^n)))] \leq d$$

(The above mappings define a code $(n, 2^{nR}, \Delta)$.) The lossy source coding with side information problem is to determine the set $\mathscr{R}$ of all achievable pairs (R, d). Further, $R^*_Y(d)$ is defined as the minimum R over all achievable pairs (R, d).

The Wyner-Ziv result for lossy source coding with side information is stated as follows: suppose the joint distribution of X and Y is given by $Q(x, y)$=Pr$\{X = x, Y = y\}$, and a distortion measure $D(x, \hat{x}) \geq 0$ is defined for the set $\mathcal{X} \times \hat{\mathcal{X}}$, then it is shown that

$$R^*_Y(d) = \min_{p(w|x)} \min_{f} (I(X; W) - I(Y; W)) \tag{1}$$

where W is an auxiliary random variable satisfying

$$\text{(a)} \sum_{w \in \mathcal{W}} p(x, y, w) = Q(x, y), \quad \text{(b)} \; p(x, y, w) = Q(x, y)p_t(w|x)$$

The minimization in (1) is performed over all functions $f : \mathcal{Y} \times \mathcal{W} \to \hat{\mathcal{X}}$ and conditional probability functions $p(w|x)$, $|\mathcal{W}| \leq |\mathcal{X}| + 1$, such that $E[D(x, \hat{x})] = \sum_x \sum_w \sum_y Q(x, y)p(w|x)D(x, f(y, w)) \leq d$. We will discuss the result and the proof.

4 Example of Lossy Source Coding with Side Information

Suppose the side information Y is a noisy version of X, i.e., $Y = X + N$. Such a situation can arise in practice. For instance, if an existing analog system such as analog TV needs upgrading, instead of replacing the system, a side digital channel can be added to the existing system. The decoder can use the additional bits sent in the digital channel (in addition to the noisy signal Y received from the analog channel) to obtain a better reconstruction of the signal.

The following example from [6] helps to illustrate the principle of source coding with side information and gives a better understanding of the result in (1). Let X and Y be continuous valued random processes that are described by i.i.d. sequences $\{X_i\}$ and $\{Y_i\}$ respectively. Let N be a zero mean

Gaussian random process independent of X and Y that is described by the sequence $\{N_i\}$. In this example, X is quantized using an 8 level scalar quantizer. First, consider the situation where no side information is available to the decoder. The Lloyd-Max algorithm partitions the real line into 8 regions $\Gamma_i, i = 0, 1, \ldots, 7$, according to the marginal distribution of X. The algorithm also assigns a point (which we call the active codeword) to each region; the active codeword is the centroid of each region, calculated from the marginal distribution of X. A sample from X denoted by x belonging to the region Γ_i is quantized to the active codeword w_i of that region. The encoder transmits the index i of the region containing the sample. The decoder estimates the original sample as:

$$\hat{x} = \arg\min_{a \in \mathbb{R}} E\Big[D(x, a) \mid X \in \Gamma_i\Big] \tag{2}$$

If the distortion measure is defined as $D(x, y) = (x - y)^2$ and if X is Gaussian, then the estimate of the decoder is equal to the active codeword w_i of region Γ_i.

Now suppose side information $Y = y$ is available to the decoder, then the decoder's estimate is:

$$\hat{x} = \arg\min_{a \in \mathbb{R}} E\Big[D(x, a) \mid X \in \Gamma_i, Y = y\Big] \tag{3}$$

Observe that this estimate must be better than the previous one in terms of reducing the distortion. We'll now see how this side information can reduce the transmission rate.

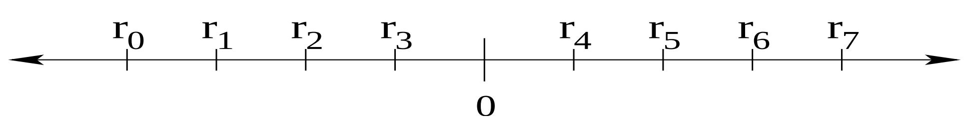

Figure 4: Scalar quantizer with 8 levels.

The 8 level scalar quantizer divides the real line into 8 regions, $\Gamma_i = (\frac{r_{i-1}+r_i}{2}, \frac{r_i+r_{i+1}}{2}]$, $(r_{-1} = -\infty, r_8 = +\infty)$. The active codewords r_i, $0 \leq i \leq 7$ are found by the Lloyd-Max algorithm as before based on the marginal distribution of X (Figure 4). Let the quantized version of X be described by the random variable W (i.e, W takes values in the set $\{r_0, r_1, \ldots, r_7\}$). The set of active codewords is divided into $M = 2$ groups $\{r_0, r_2, r_4, r_6\}$ and $\{r_1, r_3, r_5, r_7\}$. We can think of the two groups as cosets of the channel code $C = \{r_0, r_2, r_4, r_6\}$. The rate of the channel code is $R_c = 2$ bits/sample. The side information Y can be considered as the output of a fictitious channel $P(Y|W)$ having W as the input and Y as the output. The capacity of this channel is given by $I(Y; W)$.

The encoder first quantizes the sample into the active codeword w. The encoder sends the index of the coset containing w. Since there are two cosets in this example, the encoder sends 1 bit/sample. The decoder on receiving this index looks for the most likely codeword in the coset. Since the decoder has access to $Y = y$, it decodes y (or, estimates $\hat{w}$) by looking for the most likely active codeword in the coset indicated by the index[3]. There is a probability of error associated with this decoding. (This probability of error can be made very small if a channel code having a large minimum distance is chosen.) After estimating the active codeword $\hat{w}$, the decoder estimates the original sample as in (3), where Γ_i is the region containing $\hat{w}$. The encoding and decoding procedure is as shown in Figure 5.

The above design includes the following: (i) choosing a set of active codewords for a specified quantizer, (ii) choosing a (*strong*) channel code C that divides the set of active codewords into cosets of C, (iii) estimating the active codeword in the coset from the side information available to the decoder,

[3] If X, Y, N are Gaussian, the most likely active codeword in the coset is the one closest in Euclidean distance to y.

and (iv) estimating the original sample from the region containing the active codeword so as to minimize average distortion.

The above example uses sample by sample encoding and decoding. The obvious generalization is to perform block encoding and use quantizers with memory. Further, the channel code should be designed so that the code has a good minimum distance and its rate is close to $I(Y;W)$. The expression in (1) can be explained in the following way: if there was no side information available to the decoder, then the rate distortion function is given by $R(d) = I(X;W)$ since W will be the reconstructed output of the decoder. However, if side information Y is available to the decoder, then the set of active codewords can be divided into cosets of a channel code C that is designed for the fictitious channel $P(Y|W)$. The channel code can have a rate that is at most $I(Y;W)$[4] for the decoder to estimate W reliably from Y. Further, the channel code C (a subset of active codewords) that has a rate close to $I(Y;W)$ should be designed such that C divides the set of active codewords into disjoint cosets where each coset is also a valid channel code for the $P(Y|W)$ channel. The encoder then needs to send only the index of the coset containing the active codeword. For this, it requires $\sim (I(X;W) - I(Y;W))$ bits of information, which is close to what is given in (1).

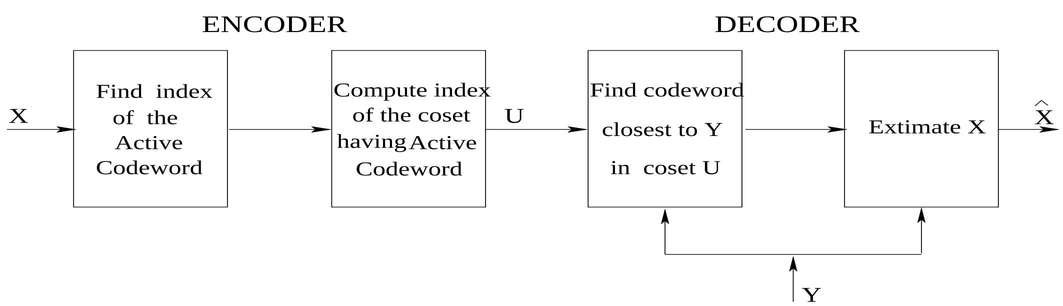

Figure 5: Wyner-Ziv coding.

5 An open problem

One long standing and unsolved problem in this area is the extension of lossy source coding with side information to two sources (and subsequently, to multiple sources). Suppose two signals X and Y are encoded separately but are decoded jointly, then the region of achievable rate pairs (R_1, R_2) for which a decoder can reconstruct the two signals X and Y within an average distortion of d, is not known. Figure 6 describes the scenario.

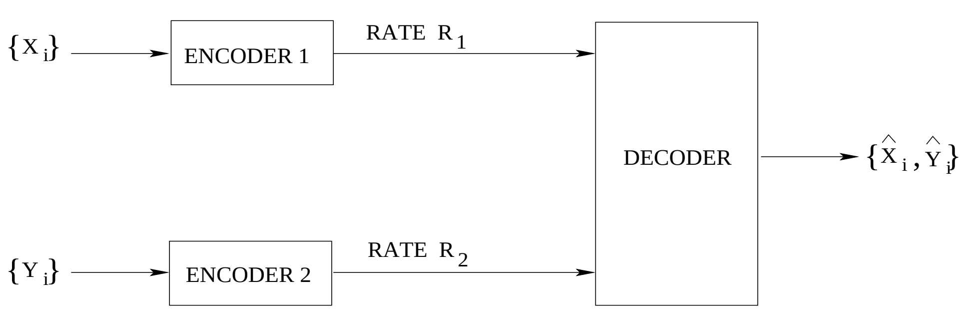

Figure 6: Rate distortion for two correlated sources.

[4] Observe that $I(Y;W)$ is the capacity of the channel described by $P(Y|W)$.

A recent paper attempted to partially solve this problem [7]. The model considered in [7] is essentially the same as Figure 6 but with additional side-information Z available to the decoder. For this case, bounds on the achievable rate region have been determined; the exact region has been found for a special case of the problem.

References

[1] A. D. Wyner, "On source coding with side information at the decoder," *IEEE Transactions on Information Theory*, May, 1975.

[2] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, Jan, 1976.

[3] S. S. Pradhan and K. Ramachandran, "Distributed source coding using syndromes (DISCUS): Design and construction," in *Proceedings of the 1999 IEEE Data Compression Conference*, (Snowbird, Utah), March 1999.

[4] A. Aaron, R. Zhang, and B. Girod, "Wyner-Ziv coding of motion video," in *Proceedings of the Asilomar Conference on Signals and Systems*, (Pacific Grove, CA), Nov 2002.

[5] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder-II: General sources," *Information and Control*, July 1978.

[6] S. S. Pradhan, *Distributed Source Coding Using Syndromes (DISCUS)*. PhD thesis, University of California, Berkeley, 2001.

[7] M. Gastpar, "The Wyner-Ziv problem with multiple sources." Submitted to IEEE Transactions on Information Theory.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley Series, 1991.